# Bubble Heap Graphs


Gi-nam Kim* — Division of Digital Media, Ajou University

Hyoji Ha† — Division of Digital Media, Ajou University

Byung-Won On‡ — Advanced Institutes of Convergence Technology, Seoul National University

Kyungwon Lee§ — Division of Digital Media, Ajou University

Manjai Lee¶ — Advanced Institutes of Convergence Technology, Seoul National University



## Abstract

In this paper, we propose **bubble heap** graphs. In general, social graphs show the overall relationship among nodes. For instance, from UN voting records, a social graph can be drawn, where each node stands for a UN member (e.g., U.S.) and further two countries show similar voting patterns if the nodes corresponding to the two countries are linked each other in the graph. In such a social graph, we can clearly figure out the overall voting patterns of all UN members. However, we often focus on one node in the graph. It is plausible that a U.S. citizen may have an interest in only U.S. and he or she wants to take a look at relationships between U.S. and each of UN members. In this case, using the existing social graph, it is hard to understand hidden insight between U.S. and any other countries. In addition, the weight value of each node is important in graph representation. This is, each country has different population and GDP. For example, GDP can be used as the weight value of nodes. The size of nodes is increased or decreased according to the weight value of each node. On this wise, given a country (called *C*) in which a user is interested, our proposed bubble heap graph effectively visualizes the relationship between *C* and each of UN members. For this bubble heap graph visualization, we present how to compute the similarity value between two nodes, and how to visualize the bubble heap graph. In particular, to prove that our proposal is general-purpose, we applied our visualization technique to two different data sets – (1) Voting records in UN General Assembly and (2) Roll call data regarding U.S. Senate sessions.


## 1 Introduction

Social networks have been widely used for visualizing information for knowledge discovery. For instance, in 2009, Andrew Odewahn visualized the U.S. Senate social network in [2]. In his work, he gathered the raw roll call data regarding the 102nd–110th Senate sessions. To generate edges in the graph, he used the traditional techniques. This is, for each bill, the vote for every senator is recorded in 'Yea', 'Nay' or 'Not Voting'. First, a similarity (*sim*) can be computed by the number of times two senators $a$ and $b$ voted the same way on the same bills. Then, if $sim(a,b) \geq$ a predetermined threshold value, the vote of $a$ is similar to that of $b$. This indicates that $a$ is connected to $b$ in the graph. He also used GraphViz's neato layout algorithm to turn an abstract representation of the nodes and edges in a graph into a picture. In particular, he focused mainly on graph visualization to paint a broad picture

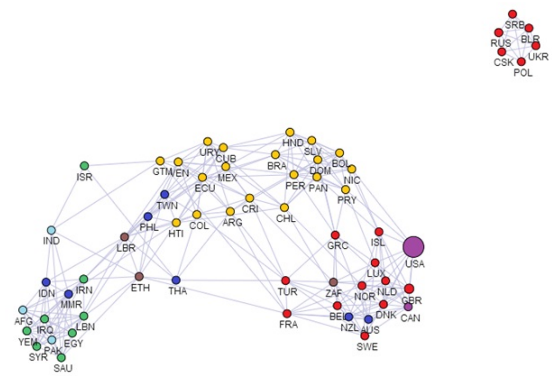



Figure 1: A social graph representing voting patterns of UN members in the bills referred between 1946 and 1950.

that revealed the structure dynamism in Senate over time. Furthermore, Pawel Bartoszek showed the voting patterns in UN General Assembly. Please take a look at the graph in [1]. He first collected the voting records of UN members through the years 2000–2008. Then, in the same way as Andrew Odewahn's approach, he created a graph in which each node stands for a country such as U.S. or Singapore. In particular, if two nodes (e.g., Korea and Japan) are linked each other in the graph, it indicates that the voting behavior of Korea is similar to that of Japan. In contrast, in case that two nodes are not connected in the graph, the countries corresponding to the two nodes show different voting patterns in UN voting records. For example, Figure 1 shows a graph in which we can see the overall voting patterns of UN members. In the graph, U.S. is linked to United Kingdom (mark as GBR). This means that the voting pattern of U.S. is close to that of United Kingdom. On the other hand, it seems that U.S. shows different voting behaviors from Russia. This is because U.S. is not linked to Russia (mark as RUS). Interestingly, there exist two disjoint graphs in Figure 1. The small graph consists of Russia, Poland, Ukraine, Serbia, and other eastern European countries. On the other hand, the large graph contains most countries around the world except Russia and eastern European countries that were part of the Communist bloc. Through the graphs in Figure 1, we can know that, for each bill, the communist countries were likely to vote similarly, whereas they voted differently, unlike U.S. and western rich countries.

In this case, a social graph (e.g., a congressional social graph or a UN General Assembly social graph) must be a useful visualization technique for understanding overall patterns of nodes' behavior (e.g., UN members' voting behaviors) from a large number of raw data (e.g., approximately 5,000 bills referred between 1946 and 2012). For visualizing such a social graph, we first compute the similarity value between nodes. Then, we create a social graph using similarity values. However, users would often like to focus on a

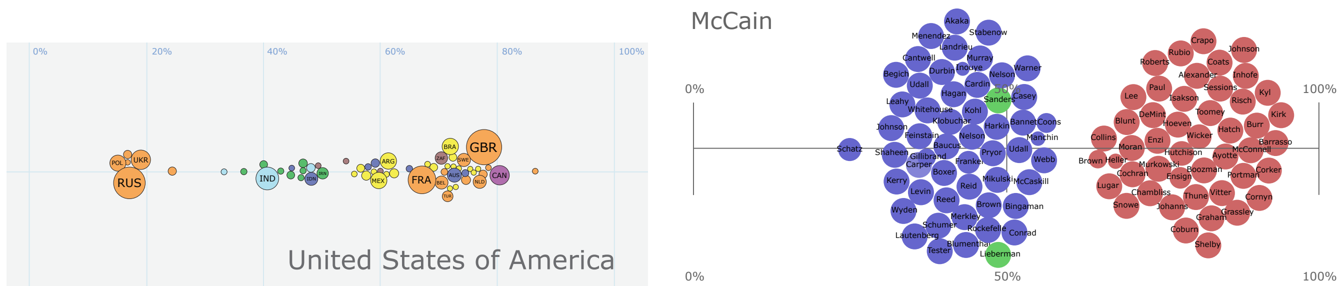

*e-mail: gnkim@ajou.ac.kr

†e-mail:hjha0508@ajou.ac.kr

‡e-mail:bwon@snu.ac.kr, Corresponding Author

§e-mail:kwlee@ajou.ac.kr

¶e-mail:manjai@snu.ac.kr




(a) A graph using UN voting records

(b) A graph using U.S. Senate roll call data

Figure 2: Bubble Heap graphs.

particular node in the graph. For instance, it is common that a U.S. citizen may have an interest in only U.S. and he or she wants to take a look at the relationship between U.S. and each of UN members. In this case, using the existing social graph, it is hard to understand hidden insight between U.S. and any other countries. In addition, the weight value of each node is important in graph representation. This is, each country has different population and GDP. For example, GDP can be used as the weight value of nodes. The size of nodes is increased or decreased according to the weight value of each node. On this wise, given a country (called $C$) in which a user is interested, our proposed bubble heap graph clearly shows the relationship between $C$ and each of UN members. Figure 2(a) depicts the social graph in Figure 1 using our proposed bubble heap visualization technique. In the bubble heap graph, the voting patterns of the communist countries such as Russia, Poland, Ukraine, etc. are considerably different from that of U.S. In addition, Figure 2(b) shows that Sen. McCain usually shows similar voting behaviors to Republicans, while his votes are different from all Democratic senators and Independents.

## 2 Main Proposal: Bubble Heap Graph

**Social Graph Formation.** In this section, we will describe our approach for computing similarity values and then visualizing bubble heap graphs. For our simple explanation, we will use U.S. Senate roll call data. The same to UN voting records. Senators are likely to select 'Yea', 'Nay', 'Abstention' or 'Not Voting' in voting bills. Based on these senators' choice, we can estimate the similarity value between votes of two senators $a$ and $b$ as follows:

$$sim(a,b) = \frac{\sum_{i \in \exists v_i(a) \wedge v_i(b)} \{c | v_i(a) = v_i(b), v_i(x) \in \{\text{Yea, Nay, Not Voting}\}\}}{\sum_{i=1}^{n} \{c | \exists v_i(a) \wedge v_i(b), v_i(x) \in \{\text{Yea, Nay, Not Voting, Abstention}\}\}} \quad (1)$$

, where $c = 1$. Note that $n$ bills are included in the Senate roll call data. In the denominator of Eq. 1, $v_i(x)$ indicates that a voter $x$ votes 'Yea', 'Nay', 'Abstention' or 'Not Voting' in the $i$-th bill, and $\exists v_i(a) \wedge v_i(b)$ denotes that both $a$ and $b$ should vote one of the four choices ('Yea', 'Nay', 'Abstention' and 'Not Voting') in the $i$-th bill. For instance, consider that $a$ votes 'Yea' and 'Not Voting' in the first and second bill. In the meantime, $b$ votes 'Yea' in the first bill. In this case, the first is the only bill voted by both $a$ and $b$. The numerator indicates how many times $a$ and $b$ vote the same way in $i \in \exists v_i(a) \wedge v_i(b)$, where $i \in n$ bills. For clear understanding, we will give a simple example as follows. Suppose that two senators $a$ and $b$ vote the five bills – $b_1$, $b_2$, $b_3$, $b_4$ and $b_5$. We also assume that $a$ votes 'Yea' in $b_1$, 'Yea' in $b_2$, 'Not Voting' in $b_3$, 'Leave' in $b_4$ and 'Nay' in $b_5$. On the other hand, $b$ also votes 'Yea' in $b_1$, 'Nay' in $b_2$, 'Not Voting' in $b_3$, 'Abstention' in $b_4$ and 'Nay' in $b_5$. In this example, $b$ votes five bills, but $a$ votes four bills – $b_1$, $b_2$, $b_3$ and $b_5$. Note that $a$ could not vote $b_4$ because he was on leave. Moreover, $b_1$, $b_2$, $b_3$ and $b_5$ are the only bills voted by both $a$ and $b$. By definition in Eq. 1, the denominator is 4. Among the four bills, $a$ and $b$ vote the same way in the three bills – $b_1$ (i.e., 'Yea' vs. 'Yea'), $b_3$ (i.e., 'Not Voting' vs. 'Not Voting') and $b_5$ (i.e., 'Nay' vs. 'Nay') so the numerator is 3. As a result, $sim(a,b) = \frac{3}{4} = 0.75$. Given a set of voters and bills, similarity values between pairs of voters are first estimated by Eq. 1. For example, the similarity value between two senators $a$ and $b$ is $sim(a,b)$ as normalized in between 0 and 1. If $sim(a,b) = 1$, it implies that the votes of $a$ and $b$ are totally the same. On the other hand, if $sim(a,b)$ is close to 0, the vote of $a$ is not correlational with that of $b$. Finally, considering four senators $a$, $b$, $c$ and $d$, we suppose that $b$ and $c$ are considered as friends of $a$ because $sim(a,b)$ and $sim(a,c)$ are the highest scores among the other similarities $sim(a,*)$s. In the same way, we also assume that $a$ and $d$ are friends of $b$ by the two highest similarities of $sim(b,*)$s. In our approach, $a$ is connected to $b$ because $a$ and $b$ are friends each other. [1] In other words, two voters $a$ and $b$ are linked in the graph if $a$ is a friend of $b$ as well as $b$ is a friend of $a$.

**Bubble Heap Graph Formation.** The social graph in the previous section can be converted into the bubble heap graph. The algorithm consists of three steps. Note that node $i$ is located in ($x_i$, $y_i$) coordinate system in our layout algorithm. In addition, $a$ is a senator that a user selects. The user wants to see the relationship between $a$ and any other senators. Here let us suppose that $b$ is one of senators except $a$. In the first step, $x_i = sim(a,b) + C \times random$ and $y_i = random$, where $C$ is required to avoid the collision of multiple nodes [2] and $C = 0.0001$. In the second step, nodes push each other apart and edges pull related nodes together. In addition, if a node collides with another node, they also push each other apart by means of Newton's laws of motion. Finally, the motion of nodes are iterated by a cooling function.

## 3 Conclusion

In this paper, we propose a bubble heap graph visualization technique in order to effectively understand how similar a pivot node is to the other nodes in the graph. **Our demonstration is available in http://politiz.org/un/.**

## References

[1] P. Bartoszek. Voting patterns in un general assembly. http://pabamapa.com/?p=648, February 2012.

[2] J. Steele and N. Iliinsky. *Beautiful Visualization*. O'Reilly Media, Inc., first edition, 2010.

[1] Note that $a$ is not connected to $b$ if $b$ has two friends $c$ and $d$. This is because $a$ is not included in the list of $b$'s friends by the highest similarity scores.

[2] Multiple nodes with the same similarity values can be located in the same position on the display. We define it as the collision of the nodes.